A map of the united states with different colored circles

Description automatically generated

**Figure 1.** North American white clover collections used in the study. The 415 accessions were collected from 43 locations; broader geographical regions are indicated by different colors and shapes as shown below the x-axis. The pie chart next to a location shows its cyanotype composition. The size of a pie chart is proportional to the number of accessions from the location. The inset figure shows the latitudinal gradient in *Growing Degree Days (> 5*°C*)* across the sampled geographical range.

A graph of a number of different numbers

Description automatically generated with medium confidence

**Figure. 2.** Linkage disequilibrium (LD) between genomic neutral markers and the cyanogenesis loci. (A), LD decay of the mild-LD-pruned dataset (241,371 SNPs). Average LD per incremental 500 bp window was plotted. Red vertical line, 25 kbp. Blue vertical line, 100 kbp. (B), interchromosomal linkage disequilibrium of genomic neutral markers and the cyanogenesis loci. Neutral marker associations between chromosomes are shown for 3,000 randomly selected SNP pairs after excluding the SNPs linked to the cyanogenesis loci. Associations of the cyanogenesis loci (*Ac* and *Li*) SNPs were calculated from the SNPs linked to the cyanogenesis loci. The difference between the two distributions was statistically tested by Kolmogorov-Smirnov (KS) test (\*\*\*p <0.001).

(A)

A picture containing colorfulness, rectangle, screenshot, pattern

Description automatically generated

(B) (C)

A picture containing text, diagram, screenshot, map

Description automatically generated

**Figure 3.** Population structure analyses. (A), ADMIXTURE output at K = 2-5. The minimum cross validation error occurs at K = 3. Locations are ordered along the x-axis by latitude, from lowest (GFL) to highest (VBC). The ancestral composition of each location is based on all accessions of a location. (B), genetic principal component analysis (PCA). The gray dots indicate the 415 accessions; the colored dots indicate the median values for each sampling location; the state abbreviations of the sampling locations are labeled with color corresponding to the broader geographical regions in **Fig. 1**. (C), Association between PC1 and the latitude for the sampling locations.

(A) (B)

A graph of a graph of a number of dots

Description automatically generated with medium confidenceA graph of a growth rate

Description automatically generated with medium confidence

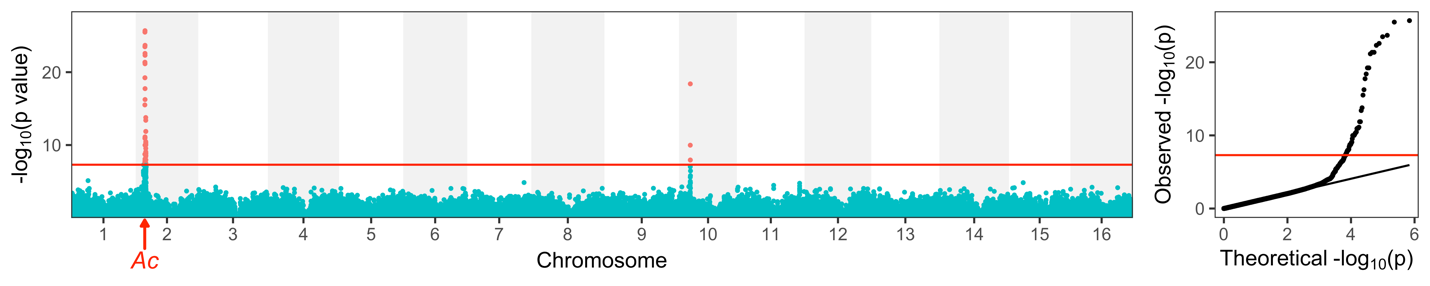
**Figure 4.** Geographical and environmental correlations with genomic differentiation for the sampled locations (n = 43). (A), Pairwise linearized genetic differentiation between locations (FST/1 – FST) is significantly associated with geographical distance and (B), Euclidean distance of *Growing Degree Days (> 5*°C*).*

(A) (B) (C)

A graph of growth in a number of degrees

Description automatically generated with medium confidence

(D)

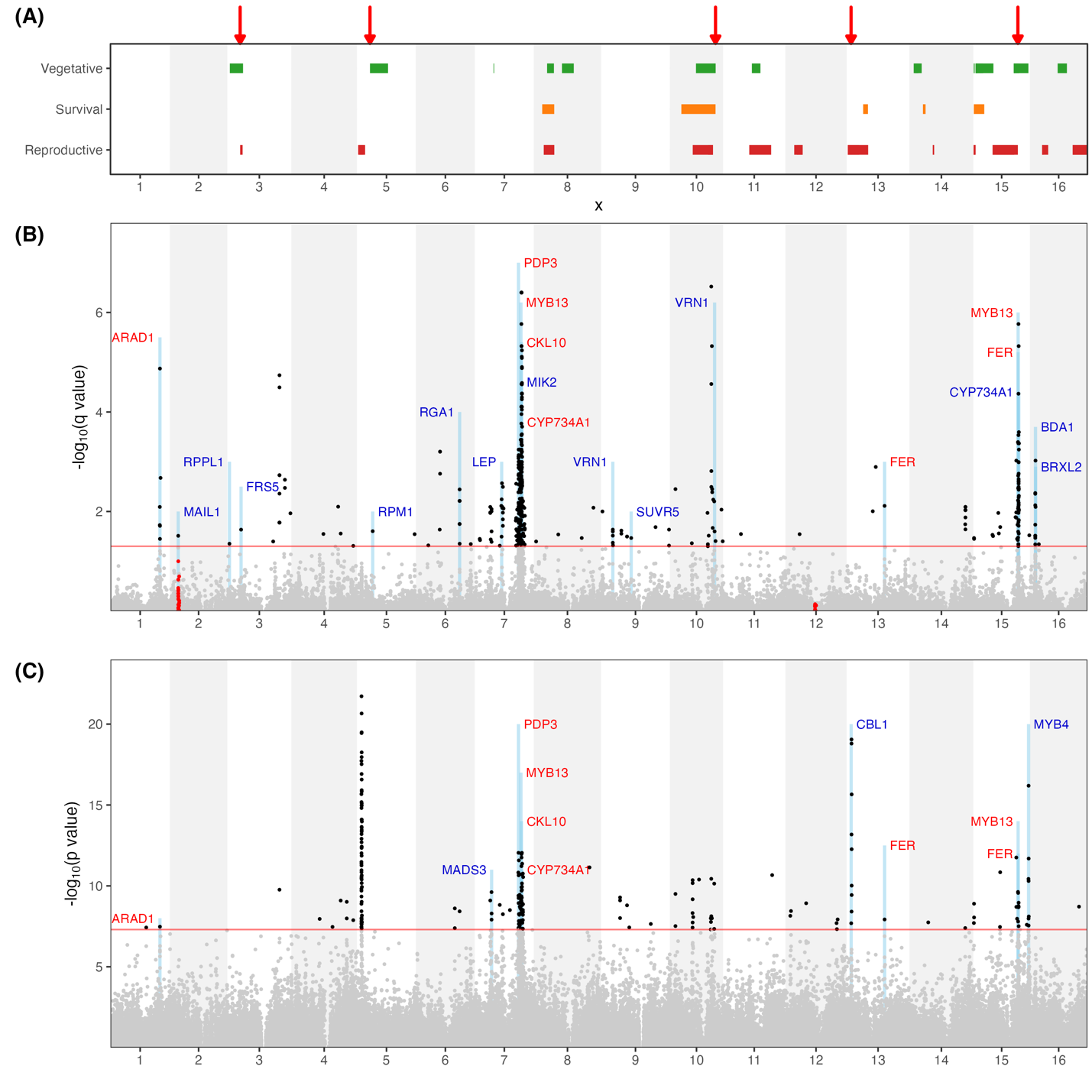


(E)

A graph with numbers and lines

Description automatically generated

**Figure 5.** Cyanogenesis polymorphism associations with environmental and genomic variation. (A-B), Associations between the copy number variation (CNV) of the cyanogenesis genes (*Ac*, *Li*) and Growing Degree Days (> 5°C). (C), observed and expected counts of the four cyanotypes among the 415 accessions. (D), GWAS of the Ac phenotype (cyanogenic glucoside) (E), GWAS of the Li phenotype (linamarase). Chromosomes 1-8 and 9-16 correspond to the two subgenomes derived from white clover’s two diploid progenitors. The red lines show the *p* value threshold at 5×10-8 Red arrows indicate the genomic locations of the *Ac* and *Li* loci.



**Figure 6.** Genome-wide scans for selection and relationship to previously identified fitness QTLs. (A), QTL mapping of the fitness traits from a reciprocal common garden experiment (Wright *et al.*, 2022). QTL intervals correspond to the 1-LOD drop. The traits are colored by fitness categories. The arrows on the top indicate that the QTL intervals overlap with the candidate genes detected in the LFMM or the PCadapt analyses. (B), LFMM test for association with *Growing Degree Days (> 5°C)*. The confounding effect of the population structure was controlled by setting the latent factor k = 3. SNPs linked to the *Ac* and *Li* loci are colored in red (Chr. 2 and 12, respectively). (C), PCadapt for genetic differentiation outlier scan. The first two PCs were included in the analysis. The *Ac* and *Li* SNPs were not labeled because PCadapt used a LD pruned dataset, which is different from the one used in GWAS. The candidate genes (related to light response, developmental process, stress or pathogen response) within 25 Kbp of the outlier SNPs are labeled. If the genes are shown in both LFMM and PCadapt results, the color is red (otherwise, blue). Note that some genes are duplicated at the proximal physical locations but only labeled once, including *RPPL1* on Chr. 3, *VRN1* on Chr. 9, *CYP734A1* on Chr. 15, and *BDA1* on Chr. 16. See **Table S3** for detailed candidate gene information.